



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Correlation Processing Of Local Seismic Data: Applications for Autonomous Sensor Deployments

D. A. Dodge

November 17, 2010

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Correlation Processing Of Local Seismic Data: Applications for Autonomous Sensor Deployments

Doug Dodge

L-046

Lawrence Livermore National Laboratory

P.O. Box 808, Livermore, CA 94551-0808

dodge1@llnl.gov

11/15/2010

Abstract

Excavation and operation of an underground facility is likely to produce an extensive suite of seismic signals observable at the surface for perhaps several km. Probably a large fraction of such signals will be correlated, so the design of a monitoring framework should include consideration of a correlation processing capability. Correlation detectors have been shown to be significantly more sensitive than beam-forming power detectors. Although correlation detectors have a limited detection footprint, they can be generalized into multi-rank subspace detectors which are sensitive over a much larger range of source mechanisms and positions. Production of subspace detectors can be automated, so their use in an autonomous framework may be contemplated. Waveform correlation also can be used to produce very high precision phase picks which may be jointly inverted to simultaneously relocate groups of events. The relative precision of the resulting hypocenters is sufficient to visualize structural detail at a scale of less than a few tens of meters.

Three possible correlation processor systems are presented. All use a subspace signal detection framework. The simplest system uses a single-component sensor and is capable of detection and classification of signals. The most complicated system uses many sensors deployed around the facility, and is capable of detection, classification, and high-precision source location. Data from a deep underground mine are presented to demonstrate the applicability of correlation processing to monitoring an underground facility. Although the source region covers an area of about 600m by 580m, all but two of the events form clusters at a threshold of 0.7. All the events could have been detected and classified by the subspace detection framework, and high-precision picks can be computed for all cluster members.

Introduction

This paper describes the use of correlation processing as a component in a very local (source-station separation $< 10\text{km}$) seismic monitoring system intended to monitor an underground facility in which repeating seismic events occur. The repeating events are not required to be co-located, but must form one or more clusters in which pairs of events are close enough to produce similar signals. For example, the process of tunneling using a drill-and-blast methodology could likely be detected and tracked by means of correlation processing. Assuming that such tracking has occurred, any subsequent explosions in cavities produced by the mining operation could be located and classified if they were large enough to be detected.

Waveform cross correlation is applicable to this problem both as a means of constructing highly sensitive detectors and as a component in producing very precise relative locations that can be used to outline underground structures produced by blasting. The next sections provide an overview of these topics

Background

It has long been recognized that in collections of data recorded by the same instrument from sources in similar locations there will be many similar seismograms (Geller and Mueller, 1980, Poupinet et al, 1984). Geller and Mueller (1980) attribute the similarity to the fact that for small magnitude earthquakes, low-pass-filtered seismograms are essentially the Green's functions, so repeated ruptures of the same place on a fault should produce the same seismogram.

Although researchers have exploited the phenomenon in a variety of different ways, most work has centered on producing high-precision relocations of clustered seismicity using correlation-derived, high-precision relative picks. For example Fremont and Malone (1987) and Got et al (1994) imaged structures underneath active volcanoes by relocation of multiplets. Rubin et al (1999) imaged seismicity on creeping sections of the Hayward fault. Hauksson and Shearer (2005) relocated 327,000 Southern California earthquakes using waveform cross correlation.

Multiplets are not restricted to be tectonic events. For example, Harris, 1991 detected quarry blasts from a range of 150 km using correlation detectors. Spottiswoode and Milev (1998) used relocated multiplets to identify planes of mining-induced seismic events. Abdul-Wahed et al (2006) developed a system for relocating mining seismicity using databases of template waveforms.

Even at very short source-receiver distances waveform correlation proves to be a useful tool. For example, Lees (1998) identified numerous multiplets in the Coso Geothermal field. Moriya et al, (2003) relocated multiplets at the Soultz geothermal field using waveforms recorded by down-hole seismometers that were apparently within a kilometer of the sources.

Correlation detectors

The most effective way to detect a known signal embedded in a time series is by means of a correlation detector (Anstey, 1966). In such a detector, the signal template is correlated with the continuous data stream and at any sample where the correlation coefficient is sufficiently high, a detection is declared. Note that such detections are also classifications. The source that produced the detection must be substantially similar in location and mechanism to the source used to create the template.

Gibbons and Ringdal (2006) demonstrated that when correlations were stacked over an array or network, there was also a significant array gain. With this technique it was possible to easily detect an aftershock of a Kara Sea event that was either undetectable or barely detectable using standard beam forming techniques. In fact, a single channel correlation detector was shown to detect signals 0.7 magnitude units lower than was possible on the array beam and using correlation detectors over the entire array pushed the detection limit down by another 0.4 magnitude units.

However, although correlation detectors are very sensitive, they are necessarily restricted in what they are able to detect. If the mechanism of a candidate source is too different from the template source or if the sources are separated by more than a wavelength or so, the candidate will not be detected. For this reason, correlation detectors are not widely used.

Subspace detectors

Seismic subspace detectors (Harris, 2006) are a generalization of correlation detectors. A subspace detector projects the data onto a vector subspace produced from a collection of seismograms in a volume of interest. The variations between the input seismograms may be due to slight differences in position or to differences in mechanism. Regardless, if the collection used to create the subspace detector is sufficiently representative, then any source in the volume can be detected with the subspace detector. The basis for the subspace is produced by aligning the windowed input seismograms, computing the singular value decomposition, and retaining a sufficient number of the left singular vectors to be able to reproduce to the required degree any of the input seismograms. Note that a subspace detector of rank 1 is exactly a correlation detector.

Because multi-rank subspace detectors have more degrees of freedom than a correlation detector they can detect a wider range of signals. Harris and Dodge (in press) used 184 automatically-generated subspace detectors to detect 676 aftershocks of the 2003 San Simeon earthquake. Many of these detectors detected only one or two events, but some of the detectors had dozens of detections associated with them. Clusters corresponding to the detection groups were identified by association with events in the ANSS bulletin. The size of the clusters identified in this manner ranged from about 2.5 km to 6.5 km in diameter. These are much larger footprints than would likely be achieved by correlation detectors.

Production of correlation picks

Poupinet *et al.* (1984) demonstrated that waveform cross correlation can be used to produce very precise (sub-sample) relative phase arrival picks. These picks can then be used to produce very high-precision relative locations of the source events.

There are two common techniques for computing the time offset between two seismograms. In the first, a line is fit to the phase of the cross-spectrum, and the slope of the line gives the delay (e.g., Poupinet *et al.*, 1984, 1984; Got *et al.*, 1994). Since the slope can be expressed to arbitrary precision, the delay may also be expressed to a precision better than the sample interval. In the second technique (Deichmann and Garcia-Fernandez, 1992) a functional form is fit to a few points surrounding the maximum of the time domain cross correlation function, and the maximum of the curve is determined analytically.

The obtainable precision of correlation picks is independent of the sample rate as long as the sample rate is high enough to represent the dominant frequencies of the data. Cheng *et al.* (2007) derived a Cramer-Rao lower bound on the precision on the precision of a correlation pick as:

$$\Delta t_{rms} = \frac{1}{\sqrt{2}\omega_0 SNR}$$

In this formulation, ω_0 is the dominant angular frequency of the phase being measured. As an example, for 30 Hz and SNR = 10, this yields a lower bound on the pick precision of ~0.0004 s. Empirically, even with data sampled at 100 Hz it is common to achieve pick precisions on the order of 1 millisecond.

Correlation picks are also more accurate (less biased) than first-break picks. For a set of similar seismograms, correlation-derived relative picks are more accurate than picks made by inspecting first

breaks because the alignment is controlled by the high-amplitude parts of the signals being correlated and is thus insensitive to the details of how the signal emerges from the noise. Dodge (1996) showed that even for signals with impulsive onsets recorded by the Northern California Seismic Network, correlation picks could improve accuracy by as much as 0.03 seconds. For signals with weak onsets, accuracy could be improved in some cases by as much as several tenths of a second.

For a group of N phase windows it is possible to compute $N(N-1)/2$ cross correlation delays. Ideally these would all be consistent; e.g. $\Delta t_{12} + \Delta t_{23} = \Delta t_{13}$. However, because of noise and other factors this is never true exactly. Analysis of the discrepancies can be performed to produce an empirical estimate of the pick precision (Cheng *et al.*, 2007), but to relocate all N events it is useful to rationalize all the delays into a consistent set. This can be done using an algorithm by Vandecar and Crosson (1990) that produces an optimum alignment for the entire cluster. The alignment so obtained is sufficient to perform a relative location of the events within a cluster.

To constrain cluster positions relative to one-another it is necessary to make the correlation-derived picks consistent from cluster to cluster. This can be done by creating one reference pick per channel-cluster and adjusting the correlation picks for each cluster so that the reference pick is aligned with the correlation picks. There are many possible picking algorithms that could be used to produce reference picks. For reviews and discussions of performance, see Withers *et al* (1998), Leonard (2000), or Baer and Kradolfer (1987).

Relocation of clusters

A number of different techniques have been used to relocate events using waveform cross correlation. The location technique used by Fremont and Malone (1987) was essentially a master event location. It was intended to precisely relocate a set of small clusters observed from a distance much greater than the extent of the clusters. For each cluster, a master event was chosen and correlation-derived differential phase times were then determined for all other cluster members. Using the assumption that the slowness vectors in the hypocentral region were the same for all cluster members, a simple inversion produced the offsets relative to the master for each cluster member.

Got *et al* (1994) used a joint inversion technique that progressively relocated events relative to their nearest neighbors. Using this technique, they were able to relocate a multiplet containing more than 250 members and around 2 km in length.

Dodge et al. (1995) converted cross correlation derived delays into standard phase picks and then performed a JHD inversion to obtain the revised event locations and velocity model. With this approach, the inter-cluster event positions were controlled primarily by the correlation-derived picks, and the intra cluster positions were controlled by manually produced first-break picks. This technique worked well for studying sequences with extents of more than 2 km.

Waldhauser and Ellsworth (2000) developed a double-difference location algorithm that combines correlation-derived picks with first-break picks to perform relative locations of seismicity. With this algorithm, the intra-cluster relative locations are computed to the accuracy obtainable from the cross correlation picks while the inter-cluster separations are computed to the accuracy obtainable from the manual picks. The authors have successfully applied the algorithm to relocate clusters having extents of many km. Because of its efficiency and flexibility, the algorithm has become widely used.

Under certain conditions, the value of the correlation between waveforms from different events can be modeled as an exponential function; e.g. $C_{ij} = e^{-r_{ij}/\alpha}$. In this formulation, r_{ij} is the inter-event separation and α is a scale factor. Israelsson (1990) used this formulation to develop a triangulation location technique which he used to relocate members of an approximate 2km by 2km cluster.

Menke (1999) also demonstrated that it was possible to relocate events using only the numerical value of the cross correlation. Used without other information only the shape of the cluster can be recovered. The shape is recovered up to a mirror reflection and a scale factor (which is a measure of regional heterogeneity). The correlation data may also be integrated into an inversion that uses travel time data. This resolves the mirror-symmetry ambiguity and determines the scale factor as well.

This method depends on the scale factor being a scalar over the region in which the relocations are to be performed. For large regions that condition is not met. For example, Nakahara (2004) determined that the correlation distance of aftershocks of the 1995 Hyogo-Ken Nanbu earthquake varied significantly from one end of the rupture zone to the other. Despite shortcomings, the potential of the method to produce source locations using single-station data makes it worth considering for some of the configurations discussed later in this paper.

Potential implementations of autonomous correlation-based processors

In the remainder of this paper I describe 3 autonomous sensor and correlation processor configurations that could be deployed for very local (< 10km) monitoring scenarios. The processing described here is

intended to provide detection and spatial classification of shallow seismic events such as might be produced during the excavation or use of underground facilities. In the most advanced configuration, high-resolution hypocenter solutions suitable for imaging the geometry of the facility or of seismic events within the facility could be produced.

A basic assumption is that there will be multiple events and that a large fraction of them will be detectable at the surface. The events must be spaced sufficiently close to one another that the waveforms observed from neighboring events can be correlated and aligned. Lacking these conditions the system may still be able to detect and locate events, but there will be no classification and the precision of the hypocenter estimates will be degraded.

The systems are described as integrated sensor-computer packages capable of doing a large amount of processing autonomously. This would have to be the case if there is very limited connection to a remote facility capable of performing the necessary processing. If the bandwidth permits, though, most of the processing could be performed at a remote facility.

The configurations are:

1. A single-component sensor with an associated processor and data storage.
2. A three-component sensor with an associated processor and data storage.
3. A collection of three or more sensors and associated telemetry allowing transmission of time-stamped data to a central node with processing and data storage capability. This configuration can be generalized to a deployment of multiple arrays with no substantial changes to most of the software algorithms since the multi-channel correlation processing is identical for both cases.

Each configuration would be capable of detection and classification of local sources. Configuration 1 might also have some ability to determine the relative geometry of multiple sources using the Menke (1999) algorithm. Configuration 2 would have limited hypocenter location capability, and configuration 3 would allow high-precision (less than a few 10s of meters) relative hypocenter locations.

No attempt is made to address the hardware issues other than to provide a description of some of the bounds related to the computational needs of the systems. All of the configurations would use a variation on the subspace detection framework described in Harris and Dodge (in press). The framework is able to cluster events using automatically created subspace detectors.

The detection framework

Subspace detectors are clearly applicable to the problem of detecting and classifying events recorded by one or a few sensors. However, there has to be a way to automatically construct and update their templates for them to be feasible in an autonomous processor. A software framework that accomplishes this task has been described by Harris and Dodge (in press).

The framework uses a bank of detectors to examine successive blocks of data from a continuous stream. For each block of data, the system accumulates triggers from all of the detectors. Where triggers are coincident (within a certain time window) the system selects the one with the highest detection statistic and promotes it to a detection.

The framework begins processing of a data stream with just a single “boot” detector in its bank of detectors. This detector is a power detector such as an STA/LTA detector, and its detections are only used to create new correlation detectors. The boot detector has a fairly high detection threshold and its triggers are subject to a number of screening operations before they may be used for template construction. The intent is to avoid creation of correlation detectors that detect signals not from seismic events of interest. Triggers from the boot detector are only processed if no existing subspace detector has a coincident trigger.

All triggers and detections produced by the system are stored in a relational database. In the original implementation this was an Oracle database, but the current version of the framework can use an embedded database. Periodically, the system retrieves from the database a set of detections that have not yet been subject to cluster analysis. The associated waveform segments are cross correlated, and the resulting matrix of correlation coefficients is used to cluster the segments into groups. Each group is then aligned and its singular value decomposition is computed. The first N left-singular vectors whose associated singular values account for more than 90% of the energy in the eigenspectrum of the data matrix are retained. These are used to create a new rank- N subspace detector which is then added to the bank of detectors. In addition, all triggers and detections that were used in the formation of the new subspace detector are reassigned to the new detector. After the reassignment, any of the correlation detectors that now have no associated detections are removed from the detector bank. By this process, the framework gradually builds a minimal set of low-rank subspace detectors that span the entire signal space. The entire process is illustrated schematically in Figure 1.

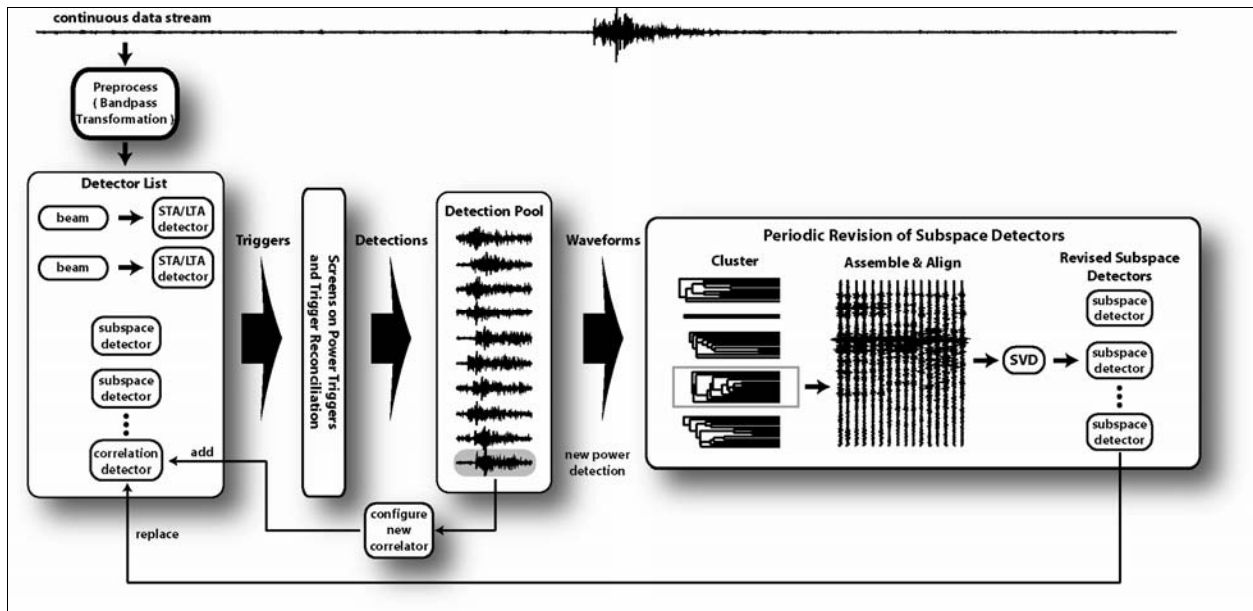


Figure 1 Block diagram of automated detection framework from Harris and Dodge (in press). The bank of detectors (left-most block) process windows from the continuous stream (shown here as a single channel). Triggers from incoherent detectors are promoted to detections when there are no co-incident subspace triggers. This results in configuration of new rank-1 subspace detector. Periodically, first-generation detections are re-clustered (right-hand block) potentially producing new multi-rank subspace detectors.

Applicability of the framework to an autonomous system

In order to be viable as part of an autonomous processor, the framework must be able to run within the memory constraints of the hardware and must be fast enough to keep up with the data flow. The current framework code is written in Java and has been run on several late-model PCs equipped with 2 or more gigabytes of RAM. In tests on the San Simeon sequence using the NVAR array, the system was able to process 10 days of data using 9 array channels and 184 detectors in less than 45 minutes. That means it was able to process at least 115,200 samples of data per second. Note that the time above includes the time spent re-clustering the detections. The high performance of the framework is due to the use of innovative signal processing techniques described in Harris and Paik (2006). By comparison, the data flow rate for a 3-component sensor sampled at 1 kHz (a possible scenario for an autonomous system) is only 3000 samples per second. So, a far less capable processor would still be able to keep up with the data flow.

Also, the algorithm implemented by the framework is parallelizable. Each detector can be run in a separate thread. This has been implemented in a revised version of the framework and tested on a dual core PC. The result was that with re-clustering turned off, the program ran in 60% of the time required

to run the single-threaded version with re-clustering turned off. (Re-clustering was turned off because that code has not been parallelized.)

The biggest demand on memory occurs during the re-clustering operation. For each group identified by the cluster method, it is necessary to compute the singular value decomposition of an M by N complex matrix where M is the length of the channel-multiplexed data vector for a single detection and N is the size of the group. There are three matrices output by the SVD. One of them (the U -matrix containing the orthonormal basis vectors) is of size M by M . As an example, to construct 5-second long templates for a 3-component sensor sampled at 1 kHz with a decimation rate of 1, M would be 15000. Stored as real-imaginary float pairs the U -matrix would require $15,000 * 15,000 * 4 * 2$ bytes (1.8 GBytes). Note: There are implementations of the SVD available that return reduced versions of the decomposition. Using such an implementation, one could arbitrarily limit the maximum rank of generated subspace detectors to a small number (e.g. 10) and control memory usage in that manner.

Also, with high decimation rates the memory requirements can be small even for the full SVD. The San Simeon data were band pass filtered from 1.0 to 3.0 Hz which allowed a decimation rate of 10. At this rate, the U -matrix in the above example would only require 18 Mbytes. For the example data set (shown later), the required bandwidth appears to be about 40 Hz, so decimation rates as high as 20 may be feasible. However, the bandwidth is likely to be location-specific, and in fact, determining an optimum bandwidth might be a required configuration process when deploying an autonomous sensor and processor.

In the current framework, processing of the data stream stops while re-clustering is occurring. That strategy works only because the data have been pre-recorded. In a real-time system it would be necessary either to maintain a data buffer of sufficient size that it could never be filled while re-clustering was occurring or else to move the re-clustering to a separate thread of execution. Overall, the second approach is probably better, particularly on a multi-processor system. The main processing thread would still need to be suspended long enough to synchronize the detector collection, but that operation only requires milliseconds.

Configuration 1

The first configuration is the simplest to implement but is also the configuration that would provide the least information. In this configuration a vertical-component seismometer would be packaged with a computer capable of running the detection framework described earlier. The system would include an

embedded database used to store configuration information, detector information including templates, trigger and detection statistics, and possibly segmented waveforms for the detections. An accurate clock would be useful, but would not be necessary unless there was a need to perform precise temporal correlations of retrieved detections with data from some other source or to use the detections in conjunction with other networks.

The system would be able to provide lists of detections grouped by detector. If an implementation of the Menke (1999) single-station location algorithm is available, the system would produce relative event locations after enough events have been detected. The usability of the locations would be dependent on the stationarity of scattering in the source region and calibration of the scale factor. Also, the absolute position and orientation of sources would be unknown with this configuration. Only the relative geometry of cluster members would be resolvable.

Configuration 2

The second configuration extends the first by replacing the vertical-component seismometer with a three-component seismometer. The detection framework would be substantially the same as in configuration 1, but the framework would be operating on channel-multiplexed data. The advantage of having three components is that it becomes possible to compute a crude source location for each detection. The location would be determined in a multi-step process. Polarization analysis of the early P-coda would be used to determine a unit-vector pointing from the receiver back to the source. Coincident with this operation, the system would estimate onset times for both P- and S-wave arrivals. With this information, and using assumed propagation velocities for P- and S-waves, the system would compute the scaled vector relating the source position to the receiver position.

There is a large body of literature on single-station location methods (See for example Frohlich and Pulliam, 1999 for a review). However, most address the domain of single-station location at regional to teleseismic distances. At the distance range considered here, the problem is slightly different. The most relevant discussion I have found is by Cichowicz (1993). In this paper, the author discusses the implementation of an in-mine system for monitoring mining-induced seismicity. The system computes the three-dimensional polarization of the P-wave, picks the P-onset, and picks the S-wave onset using a characteristic function that incorporates both changes in polarization and short-term changes in signal amplitude. The characteristics of this system would likely be applicable for the autonomous system being described here.

There is also a substantial body of literature discussing the errors inherent in azimuth estimation through polarization analysis. A theoretical analysis by Harris (1982) established the effect of noise in degrading azimuth estimates, and Rudd et al (1988) determined empirically, that stacking of like traces from 3-component stations within the NORESS array decreases azimuth estimation error. Jarpe and Dowla (1991) determined empirically that for 18 mining explosions recorded by the NORESS array, the best-case standard deviation of azimuth estimation error was ≤ 6 degrees when SNR was ≥ 5 . However, in the presence of complex local receiver structure, azimuth estimation capability can become severely degraded. Walck and Chael (1991) studied azimuth estimation performance for four RSTN stations and found that at two of the sites, receiver structure was the controlling factor in obtaining useful estimates. At station RSSD even with the best combination of parameters, 40 to 50% of the estimates had errors greater than 20 degrees.

The errors in hypocenter estimation of the configuration 2 system would be a function of signal orientation estimation, velocity model errors, and phase picking error. Although there may be significant errors in each of these quantities, because of the short length scale of the problem, estimation errors could be small relative to those associated with the regional location problem. For example, for a source at the surface and 1 km distant from the sensor, a 6 degree error in azimuth estimation results in only about 0.1 km error in epicenter location. Similarly, a 10% error in P-wave velocity (assuming a whole-space Poisson solid) yields only about 0.1 km of error in the distance estimate.

The size of the picking errors will likely be a function of SNR and local scattering behavior. However, if the sources are repeating, then cross correlation picks could be produced for each cluster member. This would allow for very precise relative S-P times within clusters. The picks that are used to tie clusters together would still be subject to the errors of the automatic picker, but there is a chance that their errors could be reduced over time by using waveform alignment to suppress noise. For example, the system could compute the SVD of each group and run the picker on the left singular vector corresponding to the maximum singular value. Those picks could then be transferred to the other group members by waveform correlation

An additional important factor in obtaining epicenter estimates with this configuration is the orientation of the sensor itself. If it cannot be deployed in such a manner that its component orientations are known, then all locations computed with this configuration would be constrained only in terms of their

distance from the sensor. It might be possible to observe the geometry of the sources relative to one another, but their absolute position and orientation would be unknown.

Configuration 3

The final configuration is the most ambitious of the three. In this configuration, three or more sensor packages would be deployed around the region of interest. These packages would communicate with a processing hub (which could be one of the sensor packages) where the detection and location calculations would be performed. This network would be used to detect, classify, and produce high-precision relative locations of clustered seismicity in the region. The requirements for this configuration are more stringent than for the configurations discussed previously. The sensor positions need to be well known. Each sensor package must have accurate (sub-millisecond) timing and a communication channel with sufficient bandwidth to allow transmission of the digitized waveform samples to a processing hub. However, only the processing hub would need to have significant computing and storage capabilities.

The principal advantage of configuration 3 relative to the other two configurations would be in its ability to produce high-precision locations (uncertainties of a few tens of meters or less). This high precision would be obtained by using correlation picks and relative location techniques.

In this configuration, the detection and clustering operation would take place at the processing hub using a multi-channel subspace detector. By processing all channels together no association would be required. Also the detector sensitivity would be improved because information about the inter-channel signal structure would be included in performing template matches. However, a practical system must be able to accommodate missing information. Not every sensor will be 100% operational during a deployment. Therefore, the framework must have a mechanism for adaptively removing template channels that are missing from the input stream.

Example data

The data in this section were collected from a mine in northern Idaho. This is a very local monitoring scenario, so applicability of some of the algorithms mentioned in this paper can be demonstrated with these data. The seismograms were recorded by a Geotech S-13 vertical-component seismometer located on the surface near the mine plant. They are in 20 second long segments sampled at 2 KHz. The events are mostly type A rock bursts (Richardson and Jordan, 2002) that were generated by mining activity along a steeply dipping ore body located about 1.3 km to the north of the seismometer. Over

the one-year period in which these events occurred, approximately 158,000 cubic meters of rock was excavated from a region extending about 600m along strike and 580 m along dip. The uppermost part of the excavation is about 1.7 km below ground level. There are 32 events in the data set.

Applicability for processing with subspace detector

Because the data are segmented, they were not processed with the detection framework. Instead, they were processed using only the clustering algorithm used by the framework. The waveforms are quite similar, and with the clustering threshold (0.3) that we used to process the San Simeon sequence, all or most of these events would have ended up in one group. Here a threshold of 0.7 was used instead. Results are shown in Figure 2. Clustering resulted in formation of 5 groups, the largest having 16 members. Three events were left un-clustered. Of the un-clustered events, one (EVID 35) is clearly not from this mine based on the S-P interval.

Although the events in this example are not explosions, they are caused by blasting and occur in close proximity to the causative blasts. This provides some confidence that similar groupings would be obtained if the causative blasts were recorded instead. The blasts do not show up in any of these records because the gain of the seismometer has been turned down to suppress noise from heavy rotating machinery near the seismometer and because the operators are not interested in collecting the blast data.

In actual operation, the framework would create 5 rank-1 subspace detectors based on the clustering of Figure 2. To illustrate the ability of multi-rank subspace detectors to match a wider range of signals, I created a rank-5 subspace basis and used it to reconstruct the data. The reconstruction is shown in Figure 3. This shows the seismograms from Figure 1 (minus the seismogram from event 35) windowed and aligned. Superimposed on each seismogram is a reconstruction produced by the rank-5 basis. Some of the reconstructions are almost perfect and all of them are highly correlated with their target seismogram. A subspace detector with the same basis could detect and classify all of these events.

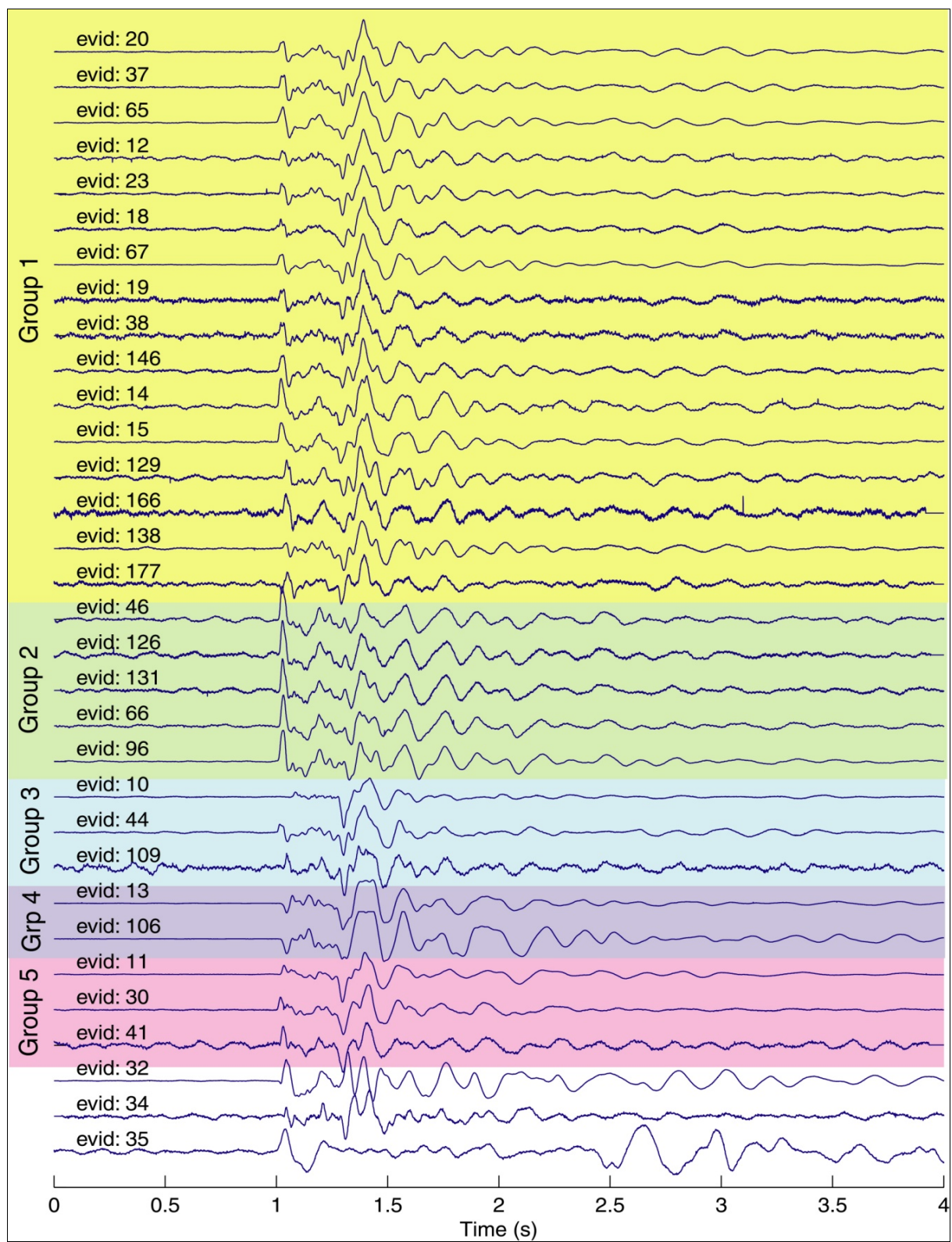


Figure 2. Waveforms from 32 seismic events after clustering. The waveforms are sorted by similarity. The color coding shows the groups determined using a complete-link cluster algorithm with threshold of 0.7.

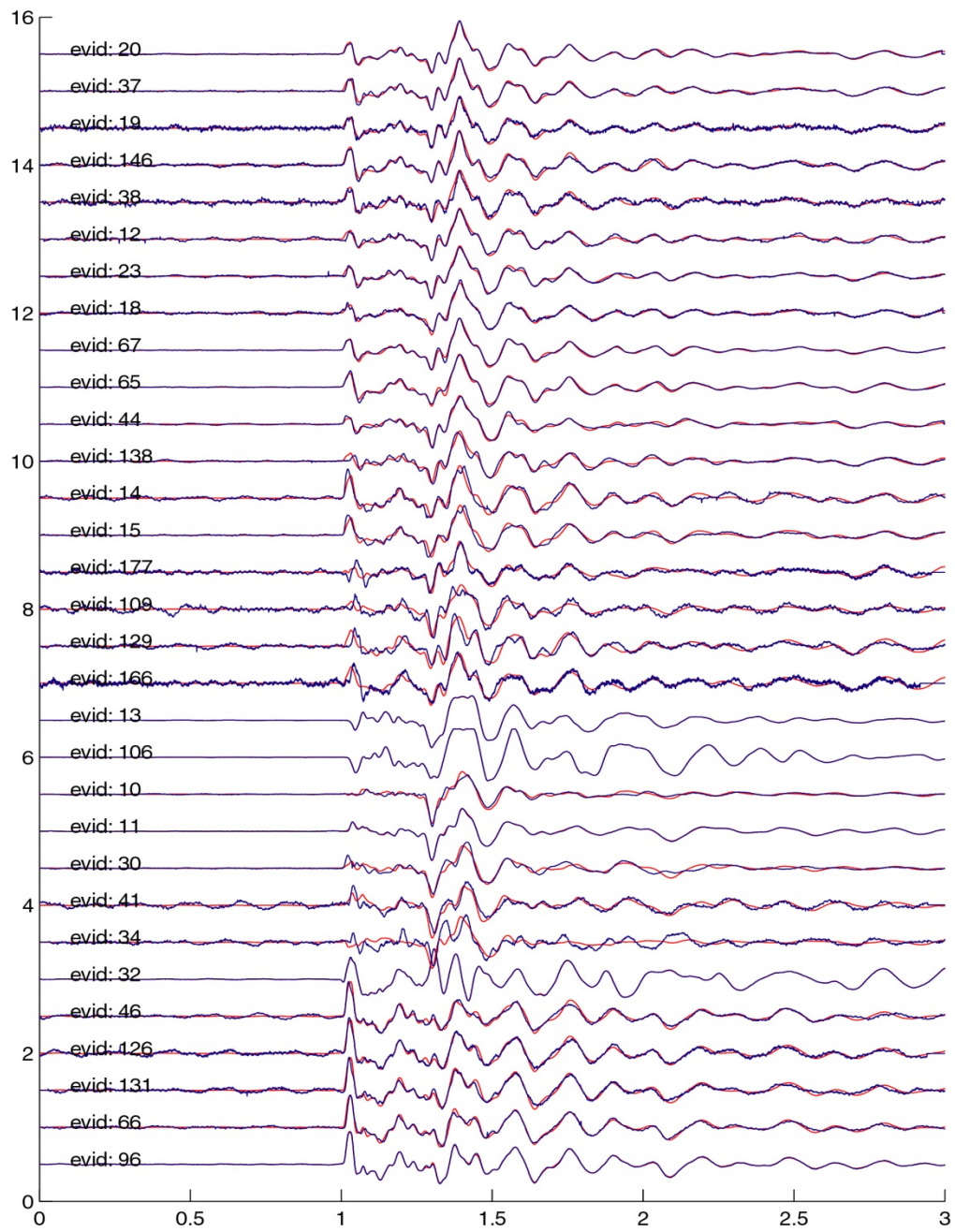


Figure 1 Reconstruction of 31 of the seismograms using the first 5 left-singular vectors from the singular value decomposition of all 31 seismograms. Blue traces are the original and red are the reconstructions.

Technology Gaps and Recommendations

All of the software systems described here are realizable. However, there is still development required to make them sufficiently robust and adaptable that they would produce reliable results in an autonomous deployment. Some of the issues are discussed below.

The subspace detection framework has so far been tested using only regional data. Certain parameters such as expected event duration and pass band are consequently tuned for those types of signals. Using the framework at very short hypocentral distances might only involve making appropriate parameter adjustments, but without trying out the framework on a representative local data set it is impossible to be sure there are no issues that might require more substantive change. Also, the framework currently has no capacity to handle data dropouts. At a minimum, it must be able to dynamically handle both the situation where there are more template channels than data channels and the inverse. This issue also affects the clustering process used in building multi-rank subspace detectors. All event segments that are to be compared and aligned must have the same channel basis. Some research will be necessary in order to determine the optimal way to achieve this.

It may be possible to include in configuration 1 some implementation of the Menke (1999) single-station correlation locator. However, significant algorithm development will be required to make this useful. The method has a limitation in that although high correlation of waveforms indicates small inter-event separation, poor correlation can be due to other factors. Experimentation will be required to determine over what range of correlation values (if any) the results are meaningful.

Also, how rapidly can the heterogeneity scale length vary within the source region before the method fails to produce useful results? How effective is it to stabilize the estimates using correlation-derived S-P interval data? Can the algorithm be stabilized by starting with a few ground-truth event locations that are used to determine appropriate scale factor (and velocities if S-P data are used)? Answering all these questions may require testing with several representative data sets.

Configuration 2 requires the estimation by means of polarization analysis of the slowness field incident on a 3-component sensor. This process is well-studied for the regional and teleseismic cases, but less has been published for the case where the source is very close to the sensor and all three components of the slowness field must be estimated. What kind of precision and error might one expect and how site-specific are those quantities. If answers cannot be found in the literature, then algorithm development and testing with multiple data sets will be required.

Configuration 2 also requires development/adaptation of a single station locator that uses wave field slowness and S-P intervals to calculate the hypocenter location. The algorithm for making the phase picks can be adapted from, e.g. Cichowicz, 1993, but the locator will probably have to be implemented from scratch. It will likely prove advantageous to implement it as a relative locator using correlation-derived differential S-P times and possibly differential slowness estimates. Its performance will also need to be validated using representative data sets.

Configuration 3 shares with the other configurations the development challenges related to the detection framework and the phase picking/refinement operations. In addition, there are some data

management issues related to hypocenter computations. The configuration would likely use a version of the double-difference locator of Waldhauser and Ellsworth (2000). However, the way in which the locator is used may have to be altered to accommodate the way the framework handles data. Between re-clustering operations events are added to existing clusters. Assuming that correlation-derived pick adjustments are being computed, the addition of a new event may cause a small change in all the other picks in the cluster. Should this trigger a complete recalculation of all the events detected by the system? Similarly, when re-clustering is performed, many picks may move to new clusters and all the picks from affected detectors may change slightly. Should this trigger a complete recalculation? Determining the best way to handle issues like this will require construction of a prototype system and testing with some appropriate data sets.

References

- Abdul-Wahed, M. K., M. Al Heib, and G. Senfaute (2006), Mining-induced seismicity: Seismic measurement using multiplet approach and numerical modeling, *International Journal of Coal Geology* 66, 137 – 147.
- Anstey, N.A. (1966), Correlation Techniques—A Review, *Can. J. Expl. Geophys.*, **2**, 55–82.
- Baer, M, and U. Kradolfer (1987), An automatic phase picker for local and teleseismic events, *Bull. Seismol. Soc. Am.* 77, 1437–1445.
- Cheng, X., F. Niu, P. G. Silver, S. Horiuchi, K. Takai, Y. Iio, and H. Ito (2007), Similar microearthquakes observed in western Nagano, Japan, and implications for rupture mechanics, *J. Geophys. Res.*, 112, B04306, doi:10.1029/2006JB004416.
- Cichowicz, A., 1993. An automatic S-phase picker. *Bull. Seismol. Soc. Am.* 83, 180–189.
- Deichmann, N., and M. Garcia-Fernandez (1992). Rupture geometry from high-precision relative hypocentre locations of microearthquake ruptures, *Geophys. J. Int.* 110, 501–517.
- Dodge, D., G. C. Beroza, and W. L. Ellsworth (1995). Evolution of the 1992 Landers, California, foreshock sequence and its implications for earthquake nucleation, *J. Geophys. Res.* 100, 9865–9880.
- Dodge, D. A. (1996). Micro-earthquake studies using cross correlation-derived hypocenters, *Ph.D. Thesis*, Stanford University.
- Fremont, M.J., and S. D. Malone (1987). High precision relative locations of earthquakes at Mount St. Helens, Washington, *J. Geophys. Res.* 92, 10,233–10,236.
- Frohlich, C., and J. Pulliam (1999), Single-station location of seismic events: a review and a plea for more research, *Physics of the Earth and Planetary Interiors* 113, 277-291.
- Geller, R. J., and C. S. Mueller, Four similar earthquakes in central California, *Geophys. Res. Lett.*, 7, 821-824, 1980.
- Gibbons, S.J. & Ringdal, F., 2006. The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.*, **165**, 149–166.
- Got, J. L., J. Frechet, and F. W. Klein (1994). Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *J. Geophys. Res.* 99, 15,375–15,386.
- Harris, D. B. (1982), Uncertainty in direction estimation: a comparison of small arrays and three-component stations, Lawrence Livermore Natl. Lab. Rep. UCID-19589-82, 30 pp., Lawrence Livermore Natl. Lab., Livermore, Calif.
- Harris, D. B. (1991), A waveform correlation method for identifying quarry explosions, *Bull. Seism. Soc. Am.*, 81(6), 2395-2418.

- Harris, D. (2006), Subspace detectors: Theory, Lawrence Livermore Natl. Lab. Rep. UCRL-TR-222758, 46 pp., Lawrence Livermore Natl. Lab., Livermore, Calif.
- Harris, D. B., and Dodge, D. A., (2011?) An Autonomous System for Grouping Events in a Developing Aftershock Sequence, *Bull. Seism. Soc. Am.* In Press.
- Hauksson, E., and P. Shearer (2005). Southern California hypocenter relocation with waveform cross-correlation, part 1: Results using the double-difference method, *Bull. Seism. Soc. Am.* 95, 896–903.
- Israelsson, H. (1990). Correlation of waveforms from closely spaced regional events, *Bull. Seismol. Soc. Am.* 80, 2177-2193.
- Leonard, M. (2000). Comparison of manual and automatic onset time picking, *Bull. Seismol. Soc. Am.* 90, 1384-1390.
- Lees, J. M. (1998). Multiplet analysis at Coso geothermal, *Bull. Seis. Soc. Am.* 88, 1127–1143.
- Menke, W. (1999). Using waveform similarity to constrain earthquake locations, *Bull. Seis. Soc. Am.* 89, 1143–1146.
- Moriya, H., H. Niitsuma, and R. Baria (2003). Multiplet-clustering analysis reveals structural details within the seismic cloud at the Soultz Geothermal Field, France, *Bull. Seism. Soc. Am.* 93, 1606–1620.
- Nakahara, H. (2004). Correlation distance of waveforms for closely located events—I. Implications of the heterogeneous structure around the source region of the 1995 Hyogo-Ken Nanbu, Japan, earthquake ($M_w = 6.9$), *Geophys. J. Int.* 157, 1255-1268.
- Poupinet, G., W. L. Ellsworth, and J. Frechet (1984). Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras fault, California, *J. Geophys. Res.* 89, 5719–5731.
- Rubin, A. M., D. Gillard, and J.-L. Got (1999). Streaks of microearthquakes along creeping faults, *Nature* 400, 635–641.
- Ruud, B.O., Husebye, E.S., Ingate, S.F., Christoffersson, A., 1988. Event location at any distance using seismic data from a single three-component station. *Bull. Seismol. Soc. Am.* 78, 308–325.
- Spottiswoode, S.M., and A. M. Milev (1998). The use of waveform similarity to define planes of mining-induced seismic events, *Tectonophysics*, 289, 51-60.
- VanDecar, J. C. and R. S. Crosson (1990), Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, *Bull. Seismol. Soc. Am.*, 80(1), 150-169.
- Walck, M.C., and E.P. Chael (1991). Optimal backazimuth estimation for three-component recordings of regional seismic events, *Bull. Seismol. Soc. Am.*, 81, 643-666.

Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California, *Bull. Seismol. Soc. Am.* 90, 1353–1368.

Withers, M., R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo (1998). A comparison of selected trigger algorithms for automated global seismic phase and event detection, *Bull. Seism. Soc. Am.* 88, 95-106.